Enforcing Safety at Runtime for Systems with Disturbances

Matthew Abate and Samuel Coogan

***Abstract*— An assured controller is one that enforces safety online by filtering a desired control input at runtime, and control barrier functions (CBFs) provide an assured controller that renders a safe subset of the statespace forward invariant. In this work, we present a problem formulation for CBF-based runtime assurance for systems with disturbances, and controllers that solve this problem must, in some way, incorporate the online computation of reachable sets. In general, computing reachable sets in the presence of disturbances is computationally costly and cannot be directly incorporated in a CBF framework. To that end, we present a particular solution to the problem, whereby reachable sets are approximated via the mixed-monotonicity property. Efficient algorithms exist for over-approximating reachable sets for mixed-monotone systems with hyperrectangles, and we show that such approximations are suitable for incorporating into a CBF-based runtime assurance framework.**

I. Introduction

Controllers whose safety guarantees are derived through the online enforcement of constraints are referred to in literature as *runtime assurance architectures* [1] or *active set invariance filters* (ASIF) [2], [3]. In this setting, system safety is posed as an invariance constraint, requiring that a system avoid some unsafe region of the statespace for all time. Specifications of this class are often used to describe real-word safety specifications due to the fact that the definition of real-world safety often is presented as the ability to avoid unsafe scenarios during deployment.

Numerous mechanisms exist for enforcing invariance constraints, and in particular, control barrier functions (CBFs) are well suited for this task. CBF-based runtime assurance architectures modify a suggested desired input at runtime to create a safe forward invariant region in the state space. This is a main idea of [4], [5] where the resulting controller is formulated as a quadratic program for systems with no disturbances, and this idea is extended in [6] to the setting with disturbances. A limitation here is the need to verify a controlled forward invariant region *a priori* and in general this region should be large; this problem can also be formulated as the search for a backup strategy with a corresponding controlled forward invariant region [7], [8]. The authors of [2], [3] present a CBF-based runtime assurance architecture, here formed via a verified backup strategy and safe region, which allows the system to leave the safe region. The method eases the problem of verifying a forward invariant region *a priori*, however, these works do not consider systems with disturbances. In this work we present a problem formulation for CBF-based runtime assurance for controlled dynamical systems with disturbances, and we present an example solution to this problem where nondeterminism in the system model is assessed via the mixed-monotonicity property.

Mixed-monotone systems are separable via a decomposition function into increasing and decreasing components and this enables the approximation of reachable sets [9]–[11] and the identification of attractive and forward invariant sets [10]; a similar approach is first pioneered in [12], and we refer the reader also to [13], [14] for fundamental results on monotone dynamical systems.

Efficient algorithms exist for over-approximating reachable sets for mixed-monotone systems with hyperrectangles, and we show that such approximations are suitable for incorporating into a CBF-based runtime assurance framework. As in [2], [3], our construction requires knowledge of a backup control strategy and a corresponding safe forward invariant region, however, the ASIF formed in this work allows the system to leave its safe region, and thus our construction does not require a large safe set *a priori*. A main assumption in our approach is that closed-loop backup dynamics are mixed-monotone with respect to a known decomposition function; large classes of systems have been shown to be mixed-monotone with respect to closed-form decomposition functions constructed from, *e.g.*, bounds on the system Jacobian matrix [15] or domains-specific knowledge [16], [17], and in some instances decomposition functions can also be solved for by computing an optimization problem [11].

In summary, the main contribution of this work are (i) we present a problem formulation for CBF-based runtime assurance for control systems with disturbances, and (ii) we present a specific solution to the problem statement, whereby the nondeterminism in the system model is assessed through mixed-monotonicity based reachability methods. The results and tools generated in this work are demonstrated though a case study[1].

II. Notation

We denote vector entries via subscript, *i.e.*, x_i for $i \in \{1, \cdots, n\}$ denotes the i^{th} entry of $x \in \mathbb{R}^n$, and we denote the *empty set* by $\varnothing := \{\}$.

This work was partially supported by the Air Force Office of Scientific Research under Award No: FA9550-19-1-0015.

M. Abate is with the School of Mechanical Engineering and the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, 30332, USA Matt.Abate@GaTech.edu.

S. Coogan is with the School of Electrical and Computer Engineering and the School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, 30332, USA Sam.Coogan@GaTech.edu.

[1]The code that accompanies this case study, and that generates the figures in this work, is publicly available through the GaTech FACTS Lab Github: https://github.com/gtfactslab/Abate_CDC2020. Certain proofs in this work are omitted and appear in an extended version of this work, available through ArXiv: https://arxiv.org/abs/2008.07019.

Given $x, y \in \mathbb{R}^n$ with $x_i \leq y_i$ for all i,

$$[x,\, y] := \{z \in \mathbb{R}^n \mid x_i \leq z_i \leq y_i \text{ for all } i\}$$

denotes the hyperrectangle with endpoints x and y, and

$$\langle\!\langle x,\, y \rangle\!\rangle := \{z \in \mathbb{R}^n \mid z_i \in \{x_i,\, y_i\} \text{ for all } i\}$$

denotes the finite set of 2^n vertices of $[x,\, y]$. We also allow $x_i \in \mathbb{R} \cup \{-\infty\}$ and $y_i \in \mathbb{R} \cup \{\infty\}$ so that $[x,\, y]$ defines an *extended hyperrectangle*, that is, a hyperrectangle with possibly infinite extent in some coordinates.

Let $(x,\, y)$ denote the vector concatenation of $x,\, y \in \mathbb{R}^n$, *i.e.*, $(x,\, y) := [x^T\, y^T]^T \in \mathbb{R}^{2n}$. Given $a = (x, y) \in \mathbb{R}^{2n}$ with $x_i \leq y_i$ for all i, we denote by $[\![a]\!]$ the hyperrectangle formed by the first and last n components of x, *i.e.*, $[\![a]\!] := [x,\, y]$, and similarly $\langle\!\langle a \rangle\!\rangle := \langle\!\langle x,\, y \rangle\!\rangle$.

III. Runtime Assurance for Nondeterministic Systems

We consider controlled dynamical systems with disturbances of the form

$$\dot{x} = f(x) + g_1(x)u + g_2(x)w \tag{1}$$

with state $x \in \mathcal{X} \subseteq \mathbb{R}^n$, control input $u \in \mathbb{R}^m$, and disturbance input $w \in \mathcal{W} \subset \mathbb{R}^n$. When the term $g_2(x)w$ is omitted from (1), we say the system is *deterministic*; otherwise, (1) is said to be *nondeterministic*. We denote by $\Phi(T;\, x,\, \mathbf{u},\, \mathbf{w})$ the state of (1) at time $T \geq 0$, when starting from an initial state $x \in \mathcal{X}$ at time 0 and evolving subject to a feedback controller $\mathbf{u} : \mathcal{X} \to \mathbb{R}^m$ and the disturbance signal $\mathbf{w} : [0,\, T] \to \mathcal{W}$. We assume that f, g_1, and g_2 are Lipschitz continuous functions and that $\mathbf{w}$ is piecewise continuous in time so that $\Phi(T;\, x,\, \mathbf{u},\, \mathbf{w})$ is unique when it exists.

The system (1) is paired with an *unsafe* subset of the system statespace $\mathcal{X}_{\mathrm{u}} \subset \mathcal{X}$. A control policy is safe if it avoids the unsafe set as formalized next.

Definition 1. A controller $\mathbf{u} : \mathcal{X} \to \mathbb{R}^m$ is *safe with respect to state* $x \in \mathcal{X}$ if $\Phi(T;\, x,\, \mathbf{u},\, \mathbf{w}) \in \mathcal{X} \setminus \mathcal{X}_{\mathrm{u}}$ for all $T \geq 0$ and for all $\mathbf{w} : [0,\, T] \to \mathcal{W}$. We extend this notation to sets so that $\mathbf{u}$ is *safe with respect to* $\mathcal{S} \subset \mathcal{X}$ if $\mathbf{u}$ is safe with respect x for all $x \in \mathcal{S}$. ■

One way to establish safety is through invariance.

Definition 2. Given a controller $\mathbf{u}$, a set $\mathcal{S} \subseteq \mathcal{X}$ is *robustly forward invariant* for (1) under $\mathbf{u}$ if $\Phi(T; x,\, \mathbf{u},\, \mathbf{w}) \in \mathcal{S}$ for all $x \in \mathcal{S}$, all $T \geq 0$ and all piecewise continuous disturbance inputs $\mathbf{w} : [0,\, T] \to \mathcal{W}$. ■

Suppose $\mathcal{S} = \{x \in \mathcal{X} \,|\, h(x) \geq 0\} \subset \mathcal{X} \setminus \mathcal{X}_{\mathrm{u}}$ for some continuously differentiable $h : \mathbb{R}^n \to \mathbb{R}$ and consider the pointwise-defined controller

$$\mathbf{u}^{\mathrm{CBF}}(x) = \underset{u \in \mathbb{R}^m}{\arg\min} ||u - \mathbf{u}^{\mathrm{d}}(x)||_2^2 \tag{2}$$

$$\text{s.t. } \frac{\partial h}{\partial x}(x)(f(x) + g_1(x)u + g_2(x)w) \geq -\alpha(h(x)) \quad \forall w \in \mathcal{W} \tag{3}$$

where $\alpha : \mathbb{R} \to \mathbb{R}$ is a given locally Lipschitz class-$\mathcal{K}$ function and $\mathbf{u}^{\mathrm{d}}(x)$ is some given controller. Provided the set of u satisfying the constraint (3) is nonempty for all x, $\mathcal{S}$ is robustly forward invariant for (1) and $\mathbf{u}^{\mathrm{CBF}}$ is safe with respect to $\mathcal{S}$. In this instance, h is said to be a *control barrier function (CBF)* for (1) as developed in [4]. Applying $\mathbf{u}^{\mathrm{CBF}}$ from (2)–(3) has added benefits beyond system safety and, in particular, $\mathbf{u}^{\mathrm{CBF}}$ will evaluate to $\mathbf{u}^{\mathrm{d}}$ whenever possible; thus, if $\mathbf{u}^{\mathrm{d}}$ has performance advantages over $\mathbf{u}$, then $\mathbf{u}^{\mathrm{CBF}}$ will retain these advantages.

It is the primary focus of this paper to design safe controllers for the system (1). To that end, we assume knowledge of a backup controller which is safe with respect to some subset of the statespace by virtue of a robustly invariant backup region as defined next.

Definition 3. The pair $(\mathbf{u}^{\mathrm{b}},\, S_{\mathrm{b}})$ with $\mathbf{u}^{\mathrm{b}} : \mathcal{X} \to \mathbb{R}^m$ and $S_{\mathrm{b}} = \{x \in \mathcal{X} \,|\, h(x) \geq 0\} \subset \mathcal{X}$ for a continuously differentiable $h : \mathcal{X} \to \mathbb{R}$ is a *backup control policy* if:

1) S_{b} is compact and $S_{\mathrm{b}} \cap \mathcal{X}_{\mathrm{u}} = \varnothing$,
2) h is concave on $\mathcal{X}$,
3) $\frac{\partial h}{\partial x} \neq 0$ on the boundary of S_{b}, and
4) there exists a class-$\mathcal{K}$ function $\alpha : \mathbb{R} \to \mathbb{R}$ such that

$$\frac{\partial h}{\partial x}(x)\left(f(x) + g_1(x)\mathbf{u}^{\mathrm{b}}(x) + g_2(x)w\right) \geq -\alpha(h(x))$$

for all $x \in S_{\mathrm{b}}$ and for all $w \in \mathcal{W}$.

In particular, the last condition above implies S_{b} is robustly forward invariant for (1) under $\mathbf{u}^{\mathrm{b}}$ via the CBF conditions discussed above and therefore $\mathbf{u}^{\mathrm{b}}$ is safe with respect to S_{b} by virtue of the first condition [6]. In this case, $\mathbf{u}^{\mathrm{b}}$ is called a *backup controller* and S_{b} its *backup region*. ■

While applying the backup controller ensures system safety, there are two primary reasons why applying such a policy is generally not preferable: (i) Backup controllers are typically designed without considering performance objectives, and (ii) $\mathbf{u}^{\mathrm{b}}$ may be safe with respect to a set larger than S_{b}, and it is possible that S_{b} is too conservative to satisfy certain performance objectives. We have already discussed how CBFs provide a solution to the first problem; however, traditional CBF based controllers are still subject to the limitations of the second problem. A solution to the second problem is presented in [2], [3] for deterministic systems, where the authors effectively increase the size of the safe region through the use of look-ahead methods.

We now have the necessary tools to define the problem of runtime assurance for nondeterministic control systems.

Problem Statement (Runtime Assurance for Nondeterministic Control Systems)**.** Assume a system of form (1) and a set of unsafe states $\mathcal{X}_{\mathrm{u}} \subset \mathcal{X}$. Additionally, assume a backup control policy $(\mathbf{u}^{\mathrm{b}},\, S_{\mathrm{b}})$, and assume a desired controller $\mathbf{u}^{\mathrm{d}}$ that satisfies some performance objective but is perhaps not safe with respect to S_{b}. The objective is to design a controller $\mathbf{u}^{\mathrm{ASIF}}$ such that $\mathbf{u}^{\mathrm{ASIF}}$ is safe with respect to S_{b} and such that $\mathbf{u}^{\mathrm{ASIF}}(x)$ evaluates to $\mathbf{u}^{\mathrm{d}}(x)$ when it is safe to do so. ■

A controller $\mathbf{u}^{\mathrm{ASIF}}$ which solves the problem statement

is referred to as an *assured controller* or an *active set invariance filter* (ASIF). We particularly aim for a solution that provides an assured controller that need not render S_b forward invariant; it may be the case, for instance, that for certain initial conditions $x \in S_\mathrm{b}$, the system (1) will be driven out of S_b by $\mathbf{u}^{\mathrm{ASIF}}$ and may not return. Nonetheless, by virtue of the fact that $\mathbf{u}^{\mathrm{ASIF}}$ is an assured controller we have that $\mathbf{u}^{\mathrm{ASIF}}$ is safe with respect to S_b.

In Section V, we present a solution to the problem statement which allows the system to leave the, perhaps conservative, safe set S_b. In our proposed solution, we specifically address nondeterminism in the system model through mixed-monotonicity based reachability methods.

IV. Preliminaries on Mixed-Monotone Systems

Before visiting the general setting of (1), we first consider the nondeterministic autonomous system

$$\dot{x} = F(x, w) \tag{4}$$

and recall fundamental results in mixed-monotonicity theory. As before, we let $\mathcal{X}$ and $\mathcal{W}$ denote the state and disturbance spaces of (4), respectively, where we now assume $\mathcal{X}$ is an extended hyperrectangle and $\mathcal{W}$ is a hyperrectangle, with $\mathcal{W} := [\underline{w}, \overline{w}]$ for $\underline{w}, \overline{w} \in \mathbb{R}^m$ and $\underline{w}_i \leq \overline{w}_i$ for all i.

Definition 4. Given a locally Lipschitz continuous function $d : \mathcal{X}\times\mathcal{W}\times\mathcal{X}\times\mathcal{W} \to \mathbb{R}^n$, the system (4) is *mixed-monotone with respect to* d if all of the following hold:

- For all $x \in \mathcal{X}$ and all $w \in \mathcal{W}$, $d(x, w, x, w) = F(x, w)$.
- For all $i, j \in \{1, \cdots, n\}$ with $i \neq j$, $\frac{\partial d_i}{\partial x_j}(x, w, \widehat{x}, \widehat{w}) \geq 0$ for all $x, \widehat{x} \in \mathcal{X}$ and all $w, \widehat{w} \in \mathcal{W}$ whenever the derivative exists.
- For all $i, j \in \{1, \cdots, n\}$, $\frac{\partial d_i}{\partial \widehat{x}_j}(x, w, \widehat{x}, \widehat{w}) \leq 0$ for all $x, \widehat{x} \in \mathcal{X}$ and all $w, \widehat{w} \in \mathcal{W}$ whenever the derivative exists.
- For all $i \in \{1, \cdots, n\}$ and all $k \in \{1, \cdots, m\}$, $\frac{\partial d_i}{\partial w_k}(x, w, \widehat{x}, \widehat{w}) \geq 0$ and $\frac{\partial d_i}{\partial \widehat{w}_k}(x, w, \widehat{x}, \widehat{w}) \leq 0$ for all $x, \widehat{x} \in \mathcal{X}$ and all $w, \widehat{w} \in \mathcal{W}$ whenever the derivative exists. ■

If (4) is mixed-monotone with respect to d, d is said to be a *decomposition function* for (4), and when d is clear from context we simply say that (4) is mixed-monotone. The mixed-monotonicity property is useful for, *e.g.*, efficient reachable set computation, and these techniques have been applied in domains including transportation system [16], and biological systems [17].

Let $\Phi^F(T; x, \mathbf{w})$ denote the state of (4) reached at time $T \geq 0$ starting from $x \in \mathcal{X}$ at time 0 under the piecewise continuous input $\mathbf{w} : [0, T] \to \mathcal{W}$, and let

$$R(T; x) := \{\Phi^F(T; x, \mathbf{w}) \in \mathcal{X} \,|\, \mathbf{w} : [0, T] \to \mathcal{W}\} \tag{5}$$

denote the time-T forward reachable set of (4) from the initial condition $x \subseteq \mathcal{X}$. We next recall how over-approximations of reachable sets can be efficiently computed by considering a deterministic auxiliary system constructed from the decomposition function.

Assume (4) is mixed-monotone with respect to d, and construct

$$\begin{bmatrix}\dot{x}\\ \dot{\widehat{x}}\end{bmatrix} = e(x, \widehat{x}) := \begin{bmatrix}d(x, \underline{w}, \widehat{x}, \overline{w})\\ d(\widehat{x}, \overline{w}, x, \underline{w})\end{bmatrix}. \tag{6}$$

The system (6) is the *embedding system* relative to d, and for $x \in \mathcal{X}$, we denote by $\Phi^e(T; x)$ the state of (6) at time $T \geq 0$ when initialized at $(x, x) \in \mathcal{X} \times \mathcal{X}$ at time 0.

Proposition 1 ([10, Prop. 1])**.** *For $x \in \mathcal{X}$, if $\Phi^e(t; x) \in \mathcal{X} \times \mathcal{X}$ for all $0 \leq t \leq T$, then $R(T; x) \subseteq [\![\Phi^e(T; x)]\!]$.* ■

V. Mixed-Monotonicity based Active Set Invariance

In this section, we present a solution to the problem statement and design a controller architecture which both allows the system to leave S_b and ensures that the system never enters $\mathcal{X}_\mathrm{u}$. The proposed controller uses a modified CBF formulation, where we now use mixed-monotonicity based reachability methods to assess the nondeterminism in the system model.

A. Problem Formulation

As prescribed in the problem statement, we assume a system of the form (1), an unsafe set $\mathcal{X}_\mathrm{u} \subset \mathcal{X}$, a backup controller $\mathbf{u}^\mathrm{b}$ with a compact backup region $S_\mathrm{b} = \{x \in \mathcal{X} \,|\, h(x) \geq 0\}$, and a desired controller $\mathbf{u}^\mathrm{d}$. As in Section IV, we assume also that $\mathcal{X}$ is an extended hyperrectangle and that $\mathcal{W} = [\underline{w}, \overline{w}]$ for given $\underline{w}, \overline{w} \in \mathbb{R}^m$ with $\underline{w}_i \leq \overline{w}_i$ for all i.

We denote by

$$\dot{x} = F^\mathrm{b}(x, w) := f(x) + g_1(x)\mathbf{u}^\mathrm{b}(x) + g_2(x)w \tag{7}$$

the closed-loop dynamics of (1) under $\mathbf{u}^\mathrm{b}$ and we let $\Phi^\mathrm{b}(T; x, \mathbf{w}) := \Phi(T; x, \mathbf{u}^\mathrm{b}, \mathbf{w})$ denote the state transition function of this system. Thus, h is a CBF for (7) and $\mathbf{u}^\mathrm{b}$ is safe with respect to S^b. Additionally, we denote by

$$S_\mathrm{b}^+(T) := \Big\{x \in \mathcal{X} \,\Big|\, \Phi^\mathrm{b}(T; x, \mathbf{w}) \in S_\mathrm{b} \text{ for all } \mathbf{w} : [0, T] \to \mathcal{W}\Big\}. \tag{8}$$

the time-T basin of attraction of S_b, which is the set of states in $\mathcal{X}$ that are guaranteed to enter S_b along trajectories of (7) within the time horizon $[0, T]$.

Remark 1. As a result of the fact that S_b is robustly forward invariant for (7), we additionally have that $S_\mathrm{b}^+(T)$ is robustly forward invariant for (7) for all $T \geq 0$. ■

As in [2], the ASIF formulation presented in this section allows the system to leave the safe set S_b in instances where the backup control policy is known to return the system to S_b on some finite time horizon. For this reason, we associate the backup control policy $(\mathbf{u}^\mathrm{b}, S_\mathrm{b})$ with a fixed backup time T_b, as formalised next.

Assumption 1. The T_b-second basin of attraction of S_b under the backup dynamics (7) does not intersect the unsafe set, *i.e.*, $S_\mathrm{b}^+(T_\mathrm{b}) \cap \mathcal{X}_\mathrm{u} = \varnothing$. ■

To verify Assumption 1 holds, one can over-approximate backward reachable sets of S_{b} under (7), and check for intersection with the unsafe set $\mathcal{X}_{\mathrm{u}}$, and many techniques allow for such an over-approximation.

Assumption 2. The backup dynamics (7) are mixed-monotone with respect to d, and we let Φ^e denote the transition function of its respective embedding system. ■

As discussed in the Introduction, Assumption 2 is not especially restrictive since large classes of systems have been shown to be mixed-monotone with closed form expressions for the decomposition function d.

B. Construction Methodology

Given $x \in \mathcal{X}$, possibly with $x \notin S^{\mathrm{b}}$, our goal is to determine a suitable value $\mathbf{u}^{\mathrm{ASIF}}(x)$; as suggested by the problem statement, $\mathbf{u}^{\mathrm{ASIF}}(x)$ should be equal or close to $\mathbf{u}^{\mathrm{d}}(x)$ if it is safe to do so. One method to determine whether or not $\mathbf{u}^{\mathrm{ASIF}}(x)$ should be equal to $\mathbf{u}^{\mathrm{d}}(x)$ is to assess the safety of the backup controller with respect to x, *i.e.*, if $R_{\mathrm{b}}(T;\, x) \subseteq S^{\mathrm{b}}$ for some $T < T_{\mathrm{b}}$ then $\mathbf{u}^{\mathrm{ASIF}}(x) = \mathbf{u}^{\mathrm{d}}(x)$ is allowed, where we let $R_{\mathrm{b}}(T;\, x)$ denote the time-T forward reachable set of (7) as in (5). We next present a family of functions that, for given $x \in \mathcal{X}$, can be used to assess whether or not $R_{\mathrm{b}}(T;\, x) \subseteq S^{\mathrm{b}}$ for some $T < T_{\mathrm{b}}$, and these functions exploit the mixed-monotonicity of (7).

Define

$$\gamma^{\mathrm{ideal}}(T;\, x) := \inf_{z \in [\![\Phi^e(T;\, x)]\!]} h(z) = \min_{z \in \langle\!\langle \Phi^e(T;\, x) \rangle\!\rangle} h(z), \quad (9)$$

where the second equality comes from the concavity on h. We show in the following lemma how γ^{ideal} is used to determine whether $x \in S_{\mathrm{b}}^{+}(T)$ for given $T \geq 0$ and $x \in \mathcal{X}$.

Lemma 1. For all $x \in \mathcal{X}$ and all $T \geq 0$, if $\gamma^{\mathrm{ideal}}(T;\, x) \geq 0$ then $x \in S_{\mathrm{b}}^{+}(T)$. ■

Next define

$$\Psi^{\mathrm{ideal}}(x) := \sup_{0 \leq \tau \leq T_{\mathrm{b}}} \gamma^{\mathrm{ideal}}(\tau;\, x). \quad (10)$$

We show in the following proposition how Ψ^{ideal} is used to assess whether $\mathbf{u}^{\mathrm{b}}$ is safe with respect to a given state.

Proposition 2. *If $\Psi^{\mathrm{ideal}}(x) \geq 0$ for some $x \in \mathcal{X}$, then applying the backup control policy starting from x at time 0 ensures that there exists a time $T \leq T_{\mathrm{b}}$ such that $R_{\mathrm{b}}(t;\, x) \subseteq S_{\mathrm{b}}$ for all $t \geq T$. In this case, we also have that $\mathbf{u}^{\mathrm{b}}$ is safe with respect to x.* ■

As a corollary to Proposition 2, note that the set

$$S_{\Psi}^{\mathrm{ideal}} := \{x \in \mathcal{X} \mid \Psi^{\mathrm{ideal}}(x) \geq 0\}$$

is robustly forward invariant on (7), and $S_{\Psi}^{\mathrm{ideal}} \subseteq S_{\mathrm{b}}^{+}(T_{\mathrm{b}})$. Thus, any controller that renders $S_{\Psi}^{\mathrm{ideal}}$ robustly forward invariant will be safe with respect to x. CBFs are well suited for this task when the relevant functions are differentiable, however, γ^{ideal} and Ψ^{ideal} are generally not differentiable due to the $\min$ construction in (9). In the next section, we present a novel soft-min construction of γ^{ideal} and Ψ^{ideal} which ensures differentiability.

C. Barrier-Based ASIF Construction

We next present a differentiable relaxation of the functions γ^{ideal} and Ψ^{ideal}, and these new functions are later incorporated in a CBF-based ASIF.

We first recall the *Log-Sum-Exponential* function.

Definition 5 (Log-Sum-Exponential)**.** We denote by

$$\mathrm{LSE}(\mathcal{S},\, p) = -\frac{1}{p} \log \sum_{s \in \mathcal{S}} \exp(-p \cdot s) \quad (11)$$

the *Log-Sum-Exponential* of the finite set $\mathcal{S} \subset \mathbb{R}$ with respect to the parameter $p > 0$. ■

The Log-Sum-Exponential has several useful properties: namely, $\mathrm{LSE}(\mathcal{S},\, p)$ is differentiable with respect to the elements of $\mathcal{S}$, and $\mathrm{LSE}(\mathcal{S},\, p)$ approximates $\min \mathcal{S}$, *i.e.*,

$$\min \mathcal{S} - \frac{n}{p} \log 2 \leq \mathrm{LSE}(\mathcal{S},\, p) < \min \mathcal{S} \quad (12)$$

for all $p > 0$, and this approximation can be made arbitrarily tight by choosing p large enough.

For fixed $p > 0$, define

$$\gamma(t;\, x) := \mathrm{LSE}(\, \{h(z) \mid z \in \langle\!\langle \Phi^e(t;\, x) \rangle\!\rangle\}\, ,\, p), \quad (13)$$

$$\Psi(x) := \sup_{0 \leq \tau \leq T_{\mathrm{b}}} \gamma(\tau, x), \quad (14)$$

and likewise $S_{\Psi} := \{x \in \mathcal{X} \mid \Psi(x) \geq 0\}$. Importantly, $\Psi(x)$ is differentiable with

$$\frac{\partial \Psi}{\partial x}(x) = \frac{\partial \gamma}{\partial x}(\tau^*(x), x) \quad (15)$$

where $\tau^*(x)$ is the maximizer from (14), *i.e.*, $\tau^*(x)$ satisfies $\Psi(x) = \gamma(\tau^*(x), x)$, and this is a result of [18, Theorem 1].

In practice, $\frac{\partial \Psi}{\partial x}(x)$ is computed as follows: First, $\Phi^e(t, x)$ is computed for t in the interval $[0,\, T_{\mathrm{b}}]$ via simulation, and the simulated trajectory is used to identify the minimizer $\tau^*(x)$ for (14). Next, $\frac{\partial \Phi^e}{\partial x}(\tau^*(x), x)$ is computed numerically; for example, n additional simulations of horizon $\tau^*(x)$ can be used to approximate the n columns of the Jacobian matrix $\frac{\partial \Phi^e}{\partial x}$. Lasty, $\frac{\partial \gamma}{\partial x}(\tau^*(x), x)$ is obtained via the chain rule using prior computations.

Lemma 2. S_{Ψ} is a strict under-approximation of $S_{\Psi}^{\mathrm{ideal}}$, *i.e.* $S_{\Psi} \subset S_{\Psi}^{\mathrm{ideal}}$. ■

As derived in Section V-B, $S_{\Psi}^{\mathrm{ideal}}$ is robustly forward invariant on (7), however, S_{Ψ} may not be. Further, S_{Ψ} may not be robustly forward invariant under any control policy, even though it is true that if $\Psi(x) \geq 0$ for some x, then applying $\mathbf{u}^{\mathrm{b}}$ will still result in eventually entering S_{b} within horizon T_{b}. This is because it is no longer the case that applying $\mathbf{u}^{\mathrm{b}}$ will keep $\Psi(x)$ from decreasing sometime before x enters S_{b}. Thus, even though a natural barrier-function-based reasoning might lead one to choose an input such that

$$\frac{d\Psi}{dt}(x(t)) \geq -\alpha(\Psi(x(t))) \quad (16)$$

for some class-$\mathcal{K}$ function $\alpha : \mathbb{R} \to \mathbb{R}$ for all time, this may not be possible when $\Psi(x)$ is close to zero, and in particular,

Algorithm 1 Runtime Assurance for Nondeterministic Control Systems

input : Desired control policy $\mathbf{u}^{\mathrm{d}} : \mathcal{X} \to \mathbb{R}^m$.
: Current State $x \in \mathcal{X}$.
: Class-$\mathcal{K}$ function $\alpha : \mathbb{R} \to \mathbb{R}$.
output: Assured control input $\mathbf{u}^{\mathrm{ASIF}}(x) \in \mathbb{R}^m$.

function $\mathbf{u}^{\mathrm{ASIF}}(x)$ =ASIF($\mathbf{u}^{\mathrm{d}}$, x, α)
  Compute:
  $u^* = \arg\min_{u \in \mathbb{R}^m} ||u - \mathbf{u}^{\mathrm{d}}(x)||_2^2$
  s.t. $\frac{\partial \Psi}{\partial x}(x)(f(x) + g_1(x)u + g_2(x)w) \geq -\alpha(\Psi(x))$
    $\forall w \in \langle\!\langle \underline{w}, \overline{w} \rangle\!\rangle$
  if Program feasible **then return** u^*
  else return $\mathbf{u}^{\mathrm{b}}(x)$

it may be the case that choosing $\mathbf{u}^{\mathrm{b}}$ violates (16). However, due to the fact that $S_\Psi \subset S_\Psi^{\mathrm{ideal}}$, if for some $x \in S_\Psi$ we have that $\mathbf{u}^{\mathrm{b}}$ violates (16), then $\mathbf{u}^{\mathrm{b}}$ is safe with respect to x from Proposition 2, and thus it is acceptable to immediately switch to the backup control policy to retain safety.

We next present our main result: an assured controller for nondeterministic control systems of the form (1). This controller is presented in pseudocode (see Algorithm 1) and control actions are chosen point-wise in time.

Let $\Phi^{\mathrm{ASIF}}(T;\, x,\, \mathbf{w})$ denote the state of (1) at time $T \geq 0$ when inputs are chosen using Algorithm 1 and when beginning from initial state $x \in \mathcal{X}$ at time 0 and when subjected to the piecewise continuous input $\mathbf{w}$.

Theorem 1. *For all initial conditions $x \in S_{\mathrm{b}}$ and any Lipschitz continuous controller $\mathbf{u}^{\mathrm{d}} : \mathcal{X} \to \mathbb{R}^m$, the controller $\mathbf{u}^{\mathrm{ASIF}}$ from Algorithm 1 is such that $\Phi^{\mathrm{ASIF}}(T;\, x,\, \mathbf{w}) \notin \mathcal{X}_{\mathrm{u}}$ for all $T \geq 0$.* ■

In summary, the assured controller $\mathbf{u}^{\mathrm{ASIF}}$ defined by Algorithm 1 (i) evaluates to the desired control input whenever possible, (ii) allows the system (1) to leave the safe region S_{b}, and (iii) ensures the system never enters the unsafe set $\mathcal{X}_{\mathrm{u}}$. Moreover, the optimization problem posed in Algorithm 1 contains only a finite number of affine constraints and, thus, the proposed assured controller can be computationally amenable to real-world applications.

VI. Numerical Example: Enforcing Inter-agent Distance Constraints on a Vehicle Platoon

Consider a platoon of 3 vehicles, whose velocity dynamics are given as

$$\dot{x}_i = \beta x_i + a_i + w_i, \tag{17}$$

where $x_i \in \mathbb{R}$ denotes the velocity of the i^{th} vehicle, a_i denotes the acceleration of the vehicle, which is controlled, $\beta = -1$ denotes a friction coefficient and $w \in [-0.1,\, 0.1]^3$ denotes a bounded additive noise term. Control decisions are made after referencing the relative displacements of vehicles in the platoon. In particular, there are two available displacement measures

$$z_1 = x_2 - x_1 \quad \text{and} \quad z_2 = x_3 - x_2, \tag{18}$$

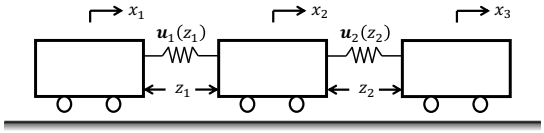


Fig. 1: Problem setting. x_1, x_2, x_3 are the vehicle velocities, and z_1, z_2 are the inter-agent distances when connectivity is given by (19). The control inputs $\mathbf{u}_1$, $\mathbf{u}_2$ effectively pull (push) the vehicles toward (away from) one another.

so that letting

$$A^T = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \tag{19}$$

we have $\dot{z} = A^T x$. The platoon dynamics then become

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \beta I & 0 \\ A^T & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} - \begin{bmatrix} A \\ 0 \end{bmatrix} \mathbf{u}(z) + \begin{bmatrix} w \\ 0 \end{bmatrix} \tag{20}$$

with control input $\mathbf{u}(z) = [\, \mathbf{u}_1(z_1),\, \mathbf{u}_2(z_2)\,]^T$. This problem setting is shown in Figure 1.

We aim to enforce inter-agent distance constraints on (20) by applying the ASIF controller presented in Algorithm 1. Specifically, we take an unsafe set

$$\mathcal{X}_{\mathrm{u}} = \{(x, z) \in \mathbb{R}^5 \mid |z_1| \geq 8 \text{ **or** } |z_2| \geq 8\}, \tag{21}$$

and we ignore vehicle collisions so that z_1 and z_2 are allowed to change sign over trajectories of (20).

We choose a backup controller

$$\mathbf{u}^{\mathrm{b}}(z) = (\, \kappa \tanh(\sigma z_1),\, \kappa \tanh(\sigma z_2)\,), \tag{22}$$

with $\kappa = 2$ and $\sigma = 1/2$. Roughly speaking, (22) acts as two identical nonlinear springs which *pull* the carts together when applied to (20); by this description, κ describes the maximum force which the springs apply before saturation, and σ describes the distance at which the springs saturate.

The closed-loop dynamics of (20) under $\mathbf{u}^b$ are mixed-monotone with respect to

$$d\left(\begin{bmatrix} x \\ z \end{bmatrix},\, w,\, \begin{bmatrix} \widehat{x} \\ \widehat{z} \end{bmatrix},\, \widehat{w} \right) = \begin{bmatrix} \beta x - A^- \mathbf{u}^{\mathrm{b}}(z) - A^+ \mathbf{u}^{\mathrm{b}}(\widehat{z}) + w \\ (A^+)^T x + (A^-)^T \widehat{x} \end{bmatrix}$$

where A^+ and A^- denote the positive and negative parts of A, respectively.

To construct a backup region S_{b}, we consider a local linearization of (7) at $(x,\, z) = 0$. The linearization is asymptotically stable to the origin and is certified by the quadratic Lyapunov function $V(x,\, z) = (x,\, z)^T P(x,\, z)$ for

$$P = \begin{bmatrix} \kappa\sigma + AA^T & -\beta A \\ -\beta A^T & (\kappa^2\sigma^2 + \beta^2) I + \kappa\sigma A^T A \end{bmatrix}. \tag{23}$$

Thus, we consider an invariant safe set

$$S_{\mathrm{b}} = \{(x,\, z) \in \mathbb{R}^5 \mid V(x,\, z) \leq \delta\} \tag{24}$$

for appropriate $\delta \geq 0$. For the parameters taken in this study, S_{b} from (24) is verified to be robustly forward invariant for the backup dynamics when $\delta = 9/4$ and, using an approach whereby backward reachable sets of S_{b} are over

approximated using the mixed-monotonicity property [10], it is additionally shown that $S_{\mathrm{b}}^{+}(T_{\mathrm{b}}) \cap \mathcal{X}_{\mathrm{u}} = \varnothing$ for $T_{\mathrm{b}} = 1$.

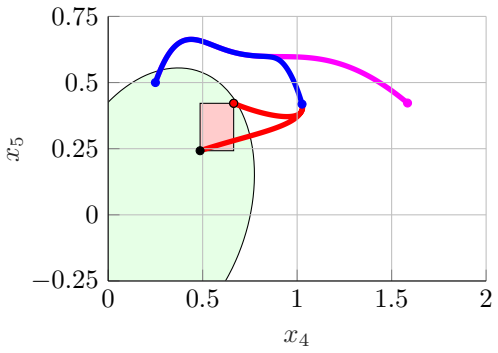
We construct an ASIF to assure the system (20), where we take the backup controller (22), safe set (24), and backup time $T_{\mathrm{b}} = 1$. In this case, γ is given by (13) where we fix $p = 1000$ and Φ^{e} is taken in reference to d. Now an assured controller is given by Algorithm 1. For the purpose of this study, we hypothesize an open-loop desired control input

$$\mathbf{u}^{\mathrm{d}}(t) = (\ -0.3\ \sin(\pi t/4),\ 0.2\cos(\pi t/2)\ ), \qquad (25)$$

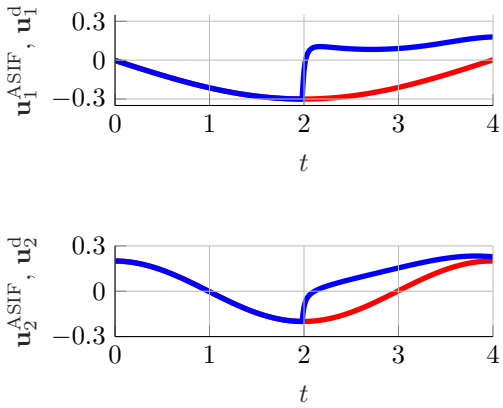
and simulate the system (20) under the ASIF controller Algorithm 1, where we let $\alpha(\psi) = 1000\psi^3$. A 4-second simulation is conducted using MATLAB 2020a and simulation results are provided in Figure 2. In the simulation the assured controller $\mathbf{u}^{\mathrm{ASIF}}$ drives the system (20) out of the safe set; however, the system remains in $S_{\mathrm{b}}^{+}(1)$ and all points along the system trajectory are safe with respect to $\mathbf{u}^{\mathrm{b}}$.

VII. Conclusion

This work presents a problem formulation for runtime assurance for control systems with disturbances, and a specific solution to the problem statement is presented, whereby the nondeterminism in the system model is accommodated via the mixed-monotonicity property. The proposed assured controller computes an optimization problem containing only a finite number of affine constraints, and we demonstrate the applicability of our construction through a case study.



(a) Cart displacement trajectory on time interval $[0,\ 4]$. The nominal trajectory $\Phi(\,\cdot\,;(x_0,\ z_0),\ \mathbf{u}^{\mathrm{d}},\ \mathbf{w})$ is shown in pink. The assured trajectory $\Phi^{\mathrm{ASIF}}(\,\cdot\,;(x_0,\ z_0),\ \mathbf{w})$ is shown in blue. Bounds on the safe backup trajectory are computed via the decomposition function d, and are shown in red. S^{b} is shown in green at time $T = 4$.



(b) Control input signals vs. time. The desired control input $\mathbf{u}^{\mathrm{d}}$ from (25) is shown in red. The applied input, which is chosen via Algorithm 1, is shown in blue.

Fig. 2: Implementing Algorithm 1 to assure (20). The carts begin with an initial velocity state $x_0 = [-1/4,\ 0,\ 1/2]^T$ and an initial displacement state $z_0 = [1/4,\ 1/2]^T$. A random disturbance $\mathbf{w} : [0,\ 4] \to \mathcal{W}$ is also chosen.

References

[1] M. Abate, E. Feron, and S. Coogan, "Monitor-based runtime assurance for temporal logic specifications," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 1997–2002, 2019.

[2] T. Gurriet, M. Mote, A. D. Ames, and E. Feron, "An online approach to active set invariance," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 3592–3599, Dec 2018.

[3] T. Gurriet, M. Mote, A. Singletary, E. Feron, and A. D. Ames, "A scalable controlled set invariance framework with practical safety guarantees," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 2046–2053, 2019.

[4] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *53rd IEEE Conference on Decision and Control*, pp. 6271–6278, Dec 2014.

[5] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," in *2019 18th European Control Conference*, pp. 3420–3431, 2019.

[6] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, pp. 3861–3876, Aug 2017.

[7] T. Schouwenaars, *Safe trajectory planning of autonomous vehicles*. PhD thesis, Massachusetts Institute of Technology, 2006.

[8] S. Bak, D. K. Chivukula, O. Adekunle, M. Sun, M. Caccamo, and L. Sha, "The system-level simplex architecture for improved real-time embedded system safety," in *2009 15th IEEE Real-Time and Embedded Technology and Applications Symposium*, pp. 99–107, IEEE, 2009.

[9] S. Coogan and M. Arcak, "Stability of traffic flow networks with a polytree topology," *Automatica*, vol. 66, pp. 246–253, Apr. 2016.

[10] M. Abate and S. Coogan, "Computing robustly forward invariant sets for mixed-monotone systems," in *2020 IEEE 59th Conference on Decision and Control (CDC)*, 2020. An extended version of this work is available through ArXiv: https://arxiv.org/abs/2003.05912.

[11] M. Abate, M. Dutreix, and S. Coogan, "Tight decomposition functions for continuous-time mixed-monotone systems with disturbances," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 139–144, 2021.

[12] G. Enciso, H. Smith, and E. Sontag, "Nonmonotone systems decomposable into monotone systems with negative feedback," *J. Differential Equations J. Differential Equations*, vol. 224, pp. 205–227, 05 2006.

[13] H. Smith, *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. Mathematical surveys and monographs, American Mathematical Society, 2008.

[14] D. Angeli and E. D. Sontag, "Monotone control systems," *IEEE Transactions on Automatic Control*, vol. 48, pp. 1684–1698, Oct 2003.

[15] P.-J. Meyer, A. Devonport, and M. Arcak, "Tira: Toolbox for interval reachability analysis," in *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '19, p. 224–229, Association for Computing Machinery, 2019. An extended version of this work appears on ArXive https://arxiv.org/abs/1902.05204.

[16] S. Coogan, M. Arcak, and A. A. Kurzhanskiy, "Mixed monotonicity of partial first-in-first-out traffic flow models," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 7611–7616, 2016.

[17] H. L. Smith, "The discrete dynamics of monotonically decomposable maps," *Journal of Mathematical Biology*, vol. 53, no. 4, p. 747, 2006.

[18] W. Hogan, "Directional derivatives for extremal-value functions with applications to the completely convex case," *Operations Research*, vol. 21, no. 1, pp. 188–209, 1973.